



**Aviation Human Factors Division
Institute of Aviation**

**University of Illinois
at Urbana-Champaign
1 Airport Road
Savoy, Illinois 61874**

**Inferring Rule-Based Strategies in
Dynamic Judgment Tasks:
Towards a Noncompensatory
Formulation of the Lens Model**

**Ling Rothrock
(Pennsylvania State University)
and
Alex Kirlik (University of Illinois)**

Technical Report AHFD-03-5/NTSC-03-1

February 2003

Prepared for

**Naval Training Systems Center
Orlando, FL**

Report Documentation Page		Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE FEB 2003	2. REPORT TYPE	3. DATES COVERED -
4. TITLE AND SUBTITLE Inferring Rule-Based Strategies in Dynamic Judgment Tasks: Towards a Noncompensatory Formulation of the Lens Model		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER	
	5e. TASK NUMBER	
	5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Training Systems Center, Orlando, FL, 32813		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT <p>Performers in time-stressed, information-rich tasks develop rule-based, simplification strategies to cope with the severe cognitive demands imposed by judgment and decision making. Linear regression modeling, proven useful for describing judgment in a wide range of static tasks, may provide misleading accounts of these heuristics. That approach assumes cue-weighting and cue-integration are well described by compensatory strategies. In contrast, evidence suggests that heuristic strategies in dynamic tasks may instead reflect rule-based, noncompensatory cue usage. We therefore present a technique, called Genetics-Based Policy Capturing (GBPC), for inferring noncompensatory, rule-based heuristics from judgment data, as an alternative to regression. In GBPC, rule-base representation and search uses a genetic algorithm, and fitting the model to data uses multi-objective optimization to maximize fit on three dimensions: a) completeness (all human judgments are represented); b) specificity (maximal concreteness); and c) parsimony (no unnecessary rules are used). GBPC is illustrated using data from the highest and lowest scoring participants in a simulated dynamic, combat information center (CIC) task. GBPC inferred rule-bases for these two performers that shed light on both skill and error. We compare the GBPC results with regression-based Lens Modeling of the same data set, and discuss how the GBPC results allowed us to interpret the high scoring performer's highly significant use of unmodeled knowledge (C=1) revealed by Lens Model analysis. The GBPC findings also allow us to now interpret a similarly high use of unmodeled knowledge (C=1) in a previously published Lens Model analysis of a different data set collected in the same experimental task. We conclude by discussing training implications, and also prospects for the development of integrated GBPC models of both human judgment and the task environment, thus providing a noncompensatory formulation of the Lens Model (a Genetics-Based Lens Model, or GBLM) of the integrated human-environment system.</p>		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 32	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

ABSTRACT

Performers in time-stressed, information-rich tasks develop rule-based, simplification strategies to cope with the severe cognitive demands imposed by judgment and decision making. Linear regression modeling, proven useful for describing judgment in a wide range of static tasks, may provide misleading accounts of these heuristics. That approach assumes cue-weighting and cue-integration are well described by compensatory strategies. In contrast, evidence suggests that heuristic strategies in dynamic tasks may instead reflect rule-based, noncompensatory cue usage. We therefore present a technique, called Genetics-Based Policy Capturing (GBPC), for inferring noncompensatory, rule-based heuristics from judgment data, as an alternative to regression. In GBPC, rule-base representation and search uses a genetic algorithm, and fitting the model to data uses multi-objective optimization to maximize fit on three dimensions: a) *completeness* (all human judgments are represented); b) *specificity* (maximal concreteness); and c) *parsimony* (no unnecessary rules are used). GBPC is illustrated using data from the highest and lowest scoring participants in a simulated dynamic, combat information center (CIC) task. GBPC inferred rule-bases for these two performers that shed light on both skill and error. We compare the GBPC results with regression-based Lens Modeling of the same data set, and discuss how the GBPC results allowed us to interpret the high scoring performer's highly significant use of unmodeled knowledge ($C=1$) revealed by Lens Model analysis. The GBPC findings also allow us to now interpret a similarly high use of unmodeled knowledge ($C=1$) in a previously published Lens Model analysis of a different data set collected in the same experimental task. We conclude by discussing training implications, and also prospects for the development of integrated GBPC models of both human judgment and the task environment, thus providing a noncompensatory formulation of the Lens Model (a Genetics-Based Lens Model, or GBLM) of the integrated human-environment system.

I. INTRODUCTION

Two-hundred and ninety people were killed on July 3, 1988 when the USS *Vincennes* mistakenly shot down an Iran Air commercial jetliner over the Persian Gulf (Fogarty, 1988). As a result of this tragedy, the U.S. Office of Naval Research established a research program on Tactical Decision Making Under Stress, or TADMUS (Collyer & Malecki, 1998). A central goal of the 7-year TADMUS program was to better understand human strengths and limitations in coping with time-stress, technological complexity, and situational ambiguity while performing judgment and decision making tasks. The TADMUS program spawned a wide range of empirical and theoretical research, characterized by close involvement among academics, government researchers, and the naval operational community. These investigators were united by a shared vision that an improved understanding of human performance in dynamic, uncertain environments could better support the design of future military training, aiding, and display systems, and thus hopefully reduce the potential for future incidence like the *Vincennes* tragedy.

The research we present in this paper was one of the many efforts initiated and supported under the TADMUS program (for a comprehensive account of TADMUS research products see Cannon-Bowers & Salas, 1998). As described in a chapter written with our colleagues in that volume (Kirlik, Fisk, Walker, & Rothrock, 1998), one of the initial steps in our own research was to visit a naval pre-commissioning team training site, consisting of a full-scale hardware and software simulation of a ship-based Combat Information Center (CIC). At this site entire CIC teams receive tactical decision making and crew coordination training just prior to taking to sea and conducting active operations. During these visits we were impressed by the tremendous amount of time and resources devoted to realism in both simulator and scenario design. At the same time, however, we were distressed by the comparatively little time and few resources devoted to providing teams with diagnostic feedback on the positive and negative aspects of their performance. Feedback given to trainees consisted of over-the-shoulder coaching, which was often disruptive and highly idiosyncratic to a particular coach's operational experiences, and team-level, classroom debriefing, which was highly abstract and delayed considerably from the training exercise itself. As a result of these observations, we made one focus of our research efforts under TADMUS to develop improved methods for performance measurement and feedback enhancement. One of the two techniques created for this purpose was a method for displaying real-time feedback to the trainee, embedded within training simulation displays, on dynamic allocation of attention to high priority events. Details on this training intervention and its empirical evaluation can be found in (Kirlik, Fisk, Walker, & Rothrock, 1998).

The purpose in the present article is to describe the second of the two feedback enhancement techniques we developed under TADMUS: a methodology for inferring, from behavioral data, the heuristic judgment strategies used by participants to cope with the time-stress and uncertainty inherent in complex, operational environments. This methodology may hold promise for advances in training technology, by making it possible to infer a performer's potential misunderstandings or oversimplifications of a judgment task from that performer's own training history. Feedback could then be conceivably be targeted toward eliminating or at least reducing these misunderstandings or oversimplifications (for a more elaborate discussion of embedded-feedback training system design, see Kozlowski, Toney, Mullins, Weissbein, Brown, & Bell, 2001).

The paper is organized in six sections. In Section II, we review the literature on how performers in time-stressed, information-rich environments cope with the cognitive demands imposed by judgment and decision tasks in order to motivate the development of our technique for inferring judgment strategies from behavioral data. In Section III, we briefly describe the historically predominant method for making such inferences, linear regression-based policy capturing, and more specifically, Brunswik's Lens Model. We conclude that section by discussing why the regression-based inferential approach may yield descriptions of judgment strategies that are inconsistent with the empirical findings described in the previous section. That discussion motivates Section IV, in which we present our noncompensatory, inferential technique based on genetic algorithms and multi-objective optimization for identifying rule-based heuristic strategies from judgment data, which we call Genetics-Based Policy Capturing (GBPC). The constructs and mathematical formalisms underlying the GBPC technique are described in detail, using a running example to illustrate each stage of model development. Section V is devoted to an empirical evaluation of the utility of the approach. Specifically, we create and compare both Lens Models and rule-based models of the same judgment data, showing how the latter helps to resolve difficulties in interpreting the regression-based representations, and also in interpreting an anomaly in a previously published Lens Modeling research using the same laboratory task (Bisantz, Kirlik, Gay, Phipps, Walker, & Fisk, 2000).

The paper concludes in Section VI with a discussion of training implications, and prospects for the development of a noncompensatory formulation of the Lens Model by using the GBPC technique to model both the human judge and the task environment. Such an approach, currently being developed, would result in a Genetics-Based Lens Model or GBLM of the entire performer-environment system. We are currently working toward a degree of formalization of the GBLM on a par with the original Lens model, and if successful, the GBLM would allow for the types of decompositions of judgment performance and analyses of adaptation enabled by the original Lens Model, but under the assumptions that both human judgment and the task environment are both well described in a rule-based, rather than linear-additive, format.

II. COPING STRATEGIES IN DYNAMIC TASKS

Human-machine systems researchers have been investigating how performers cope with time-pressure, complexity, and uncertainty in dynamic task environments since at least the 1970s (Sheridan & Johanssen, 1976). Early attempts presumed that both the task environments themselves and operator behavior could be usefully described using formal, analytical techniques from the decision sciences. As described by Klein (1999), these early attempts, based largely on prescriptive decision theory, failed to provide significant leverage for training and design, largely for two reasons. First, they were overly restrictive: human judgment and decision making in operational tasks is concerned with a wider range of phenomena than merely the crystallized moment of choice. While prescriptive decision theories focus almost exclusively on the single-shot selection of an alternative, judgment and decision making in operational settings may additionally involve situation assessment, actions taken to gather additional information, generating plausible hypotheses and alternatives, and so on. The Naturalistic Decision Making (NDM) paradigm (Klein, 1999; Klein, Orasanu, Calderwood, & Zsombok, 1993) has come to represent a broadened view of judgment and decision making, with a focus on studying "how people use their experience to make decisions in field settings" (Klein, 1999, p. 97).

NDM's focus on how people use experience touches on the second reason why prescriptive decision analysis has not proven very useful for system design and training. Prescriptive judgment and decision models provide few resources to represent the influence of experience on human behavior and performance (Orasanu & Connolly, 1993; Kirlik & Bisantz, 1999). In contrast, empirical studies of experienced performers in dynamic, uncertain environments nearly always find that a central achievement of learning is the development of "pre-established routines, heuristics, and short-cuts" (Reason, 1987, p. 468). Even those conducting empirical research within the cognitively-broadened NDM paradigm consistently find that 80 to 90% of judgments or decisions made by experienced performers are made in a rapid, intuitive process of "recognition" (Klein, 1999). Similar modes of rapid, situation-response, judgment and decision making have been found in a wide variety of dynamic, uncertain contexts, and have been described with a host of psychological constructs, including "pattern matching" (Rouse, 1983), "rule-based behavior" (Rasmussen, 1983), and "perceptual heuristics" (Kirlik, Walker, Fisk, & Nagel, 1996; Kirlik, Miller, & Jagacinski, 1993).

Despite the theoretically subtle differences in the language used to describe this type of rapid, intuitive mode of judgment and decision making, substantial evidence now exists that experienced performers in dynamic task environments will develop experiential, heuristic strategies to cope with uncertainty and time-stress. This conclusion does not imply, of course, that a performer, however experienced, will have a heuristic solution available to meet the demands imposed by every task situation. Inevitably, rare events will occur that defeat the performer's available heuristics, and may thereby either initiate a more elaborate, knowledge-based decision process (Cohen, Freeman, & Thompson, 1997; Kaempf, Klein, Thordsen, & Wolf, 1996; Rasmussen, 1983), or result in human error (Reason, 1987).

This finding, however, does not alter the fact that a majority of experienced judgments and decisions in dynamic tasks, both productive and unproductive alike, are made in a heuristic fashion. Design and training interventions must reflect this fact. Specifically, this means that task analysis, interface design, and training should focus on identifying the possibly subtle cues and situations to which a performer either does, or should, attend (Klein, 1999; Kirlik, 1995), and on how well the performer can productively use this information. The technique presented in this paper is intended to provide an additional resource to meet this need.

III. INFERRING JUDGMENT STRATEGIES FROM BEHAVIORAL DATA

A. Policy Capturing and the Lens Model

Linear regression is by far the most prevalent method of inferring possible judgment strategies from behavioral data (Hammond, 1955; Dawes & Corrigan, 1974). This judgment analysis methodology, also called "policy capturing," has been used to successfully examine a diverse set of issues including clinical judgment, conflict resolution, interpersonal learning, expertise, and the types of feedback that promote learning (for a review see Brehmer & Brehmer, 1988). Regression analysis, when applied to human judgment data, typically yields a linear-additive model of judgment. This linear model is taken to represent how a performer might weight and combine probabilistic cues in order to render a judgment or prediction about the state of the world (e.g., a physician using medical history and clinical test information to diagnose a disease).

Based on the pioneering work of the ecological psychologist Egon Brunswik (1955), an even more sophisticated and potentially useful representation of human judgment is possible when a model of the judgment environment is available (i.e., the judgment criterion can either be objectively measured or estimated by consensual expertise). This representation, called the Lens Model, is depicted in Figure 1. The Lens Model represents the judgment-environment system as a symmetrical structure. The task environment, or ecology, is represented in the left half of the figure, where the human judge is represented on the right half.

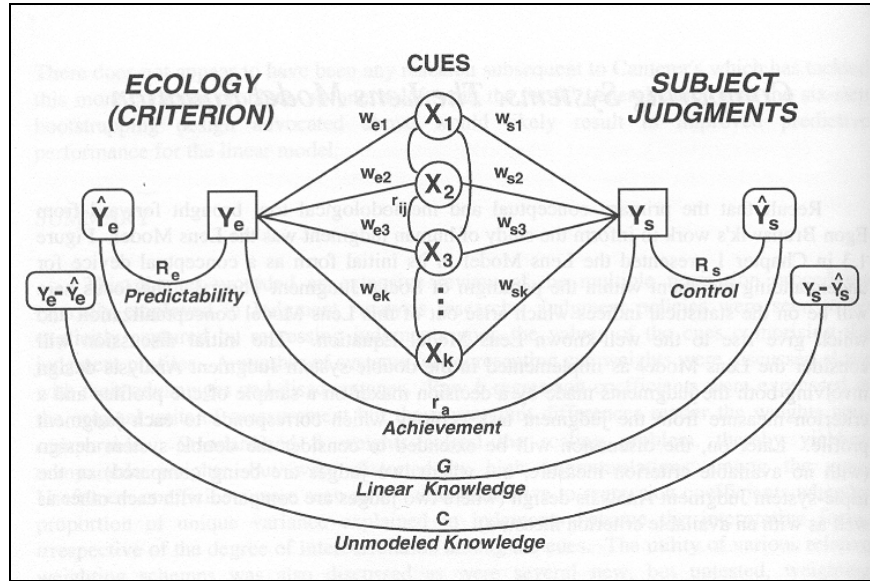


Figure 1. Lens Model with Labeled Statistical Parameters [from 24]

The symmetry inherent in this representation allows one to measure the degree of adaptation or “fit” between the judge and the demands of the judgment task. Since correlational statistics and regression were relatively new during the time in which Brunswik outlined the Lens Model, there was a need, therefore, to construct a mathematical formulation of the model that could enable efficient data analysis and modeling. The task of creating the quantitative Lens Model framework was undertaken by Hammond and his colleagues (Hammond, 1955; Hursch, Hammond, & Hursch, 1964; Tucker, 1964). The multiple linear regression model of the judge is formulated as $Y_s = \hat{Y}_s + e$ where $\hat{Y}_s = w_{s1}X_1 + w_{s2}X_2 + \dots + w_{sk}X_k$, w_{sk} are weights and e is the residual. The correlation between Y_s and \hat{Y}_s is given by R_s and represents the cognitive control (or consistency) of the judge. A corresponding multiple regression model is given for the ecology – graphically depicted as the left-hand side of the cues. For the environmental model, R_e represents the predictability of the criterion. The correlation between \hat{Y}_s and \hat{Y}_e , or G , has been labeled as linear knowledge (Hammond & Summers, 1972) to denote the linear correspondence between the judge’s decision policy and the optimal model of the criterion. The correlation between the two sets of residuals ($Y_s - \hat{Y}_s$ and $Y_e - \hat{Y}_e$), or C , is commonly called unmodeled knowledge – suggesting that if the residual variance is systematic, the judge is using a non-linear policy effectively. The remaining term, r_a , is the achievement of the judge as measured by the linear correlation between judgments and the criterion. The entire set of these

statistics are related by the Lens Model Equation (Hursch, Hammond, & Hursch, 1964; Tucker, 1964),

$$r_a = GR_e R_s + C\sqrt{(1 - R_e^2)}\sqrt{(1 - R_s^2)} \quad (1)$$

The Lens Model Equation (LME) formulates a judge's performance (or achievement) in a task in terms of components that account for linear ($GR_e R_s$) and non-linear ($C\sqrt{(1 - R_e^2)}\sqrt{(1 - R_s^2)}$) correlations.

The lens model has been applied to study problems in multiple-cue probability learning, cognitive feedback, and policy capturing (for an overview see Hammond, 1993).

Recent studies of decision making using the lens model in telerobotics (Sawaragi, Horiguchi, & Ishizuka, 2001; Horiguchi, Sawaragi, & Akashi, 2000), identification tasks in a dynamic domain (Bisantz, Kirlik, Gay, Phipps, Walker, & Fisk, 2000), and adversarial decision making (Bisantz, Llinas, & Drury, 2001) have focused on a compensatory formulation of the lens model as presented in (1). This is due to the fact that the typical method of inductive inference in lens modeling is linear regression and correlation. Therefore, while approximations to noncompensatory rules have been constructed (Einhorn, 1970; Ganzach & Czaczkes, 1995), their direct use within the lens model equation has not been investigated.

B. Potential Limitations of the Lens Model

When used to guide training and design, it is important to note that the regression approach for inferring judgment strategies from behavioral data makes specific assumptions about the cognitive processes that underlie judgment behavior. First, regression assumes that the judge has available a set of cues which he or she is able to measure. This measurement can either be binary (i.e., a cue is either absent or present), or else in terms of the magnitude of a cue value. In addition, the judge is assumed to use some form of cue weighting policy, which is correspondingly modeled by the set of weights resulting from the regression model that best fits the judge's behavioral data. Finally, a regression model assumes that the judge then integrates (the possibly differently weighted) cue values into a summary judgment. This type of weighting and summing judgment process, as represented by a linear-additive rule, has an important property: it reflects a *compensatory* strategy for integrating cue information. These strategies are compensatory in the sense that the presence of a cue with a high value, or high positive weighting can compensate for an absence of cues with moderate or low weighting. Similarly, cues with high negative weights compensate for cues with high positive weights, reflecting the manner in which a person might weigh, or trade off, evidence for and against a particular judgment.

A *noncompensatory* strategy, on the other hand, is one in which this "trading off" property is absent (Dawes, 1964; Einhorn, 1970; Gigerenzer & Goldstein, 1996). Einhorn (1970) discussed two noncompensatory judgment rules: a conjunctive rule and a disjunctive rule. A conjunctive rule describes a strategy in which every cue considered in the judgment must have a high value (or exceed some threshold) in order for the overall judgment to have high value. People being evaluated on their job performance often complain when it appears they are being

assessed by conjunctive strategy, noting that they will not receive a high evaluation or job promotion unless they perform at a high level on every dimension of evaluation. Note the noncompensatory nature of this strategy: no cue value, however highly weighted, can compensate for a low value on any one of the other cues.

A second type of noncompensatory strategy discussed by Einhorn is a disjunctive rule. A disjunctive strategy is one in which only one cue must have a high value, or exceed some threshold, in order for the overall judgment to have high value. A good example of a disjunctive strategy might be the evaluation of athletes in a professional (U.S.) football draft: a player might be highly rated if he has high value on any within the set of relevant, evaluative dimensions (e.g., speed, placekicking ability, punting ability, passing ability, etc.). Note that this strategy is noncompensatory, in the sense that a low value on a particular cue or set of cues does not detract from an overall high rating, given the presence of at least one cue with high value.

Many simple behavioral rules have a noncompensatory nature. In fact, any set of logical rules for making judgments that is inconsistent with a weighting-and-summing formula is likely to have a noncompensatory nature. In some cases, linear regression may provide an approximate fit to behavioral data generated by noncompensatory strategies such as these. However, differences may exist between the predictions of a compensatory, linear-additive model and the predictions of a noncompensatory, rule-based model in particular portions of the cue space (for a discussion and graphical depiction see Einhorn, 1970).

To take a simple example, consider a task in which the two cues are considered by a judge in an exclusive-or relationship. This example can represent the judgment of a personnel manager responsible for hiring potential job candidates. The manager looks for someone who has a degree from either one college or another. Having no degree disqualifies the candidate from insufficient credentials while having degrees from both colleges makes the candidate overqualified for the job. Therefore, one (and only one) cue must have a high value for the resulting judgment to have high value. Fitting a linear regression to data collected from such a judge results in zero weighting for the predictor coefficients. This dilemma arises because the regression finds a line that minimizes the sum of the squared errors between this line and the judgment data. Although no such line exists in an exclusive-or decision policy, the regression model forces a “compromise.”

For the purposes of the present paper, it is important to note that many if not most noncompensatory judgment strategies make lower information search and integration demands than do compensatory, linear-additive strategies. The latter always require every cue to be assessed, weighted, and combined to yield an overall judgment. Noncompensatory strategies, on the other hand, typically require fewer judgment cues to be considered, weighted, and combined to make a judgment (for a discussion, see Gigerenzer & Goldstein, 1996). Importantly, psychologists studying judgment have found that two particular task conditions are important in prompting people to shift from elaborate and exhaustive, compensatory judgment strategies, to less demanding, noncompensatory strategies to perform the same judgment task. These two conditions are task complexity (Payne, 1976), and time stress (Payne, Bettman, & Johnson, 1988; Wright, 1974). Increasing task complexity (e.g., number of cues, number of possible alternatives), and time stress both tend to increase the likelihood that people will adopt cognitively less demanding, noncompensatory strategies for making judgments.

These findings are clearly in keeping with our previous comments about the tendency of performers in complex, dynamic systems to adopt rule-based, heuristic coping strategies for handling the information processing demands of their task environments. Additionally, heuristic rule-based strategies that may initially appear to be vastly oversimplified for meeting the demands of a particular judgment task, often yield surprisingly good and robust performance when compared against much more cognitively demanding compensatory strategies (Gigerenzer & Kurz, in press; Kirlik, Walker, Fisk, & Nagel, 1996). Given that both empirical research and psychological theory strongly suggest that performers in complex, dynamic environments will develop and use noncompensatory judgment strategies, linear regression approaches for inferring these strategies from behavioral data in these contexts may be inappropriate, and may lead to misleading accounts of the behavior of these performers. As a result, a need exists to develop a technique for inferring noncompensatory, rule-based judgment heuristics from behavioral data, which does not make the compensatory assumptions underlying linear regression. The following section describes a technique developed for this purpose.

IV. INDUCTION: A NONCOMPENSATORY APPROACH

As observed by the 18th century philosopher David Hume, any knowledge derived from induction cannot, in principle, be taken as certain. While investigators have advanced the field of causality (Pearl, 2000), the preconditions of causal calculus make an application toward representative environments an unrealistic undertaking. Therefore, the purpose of induction in the context of this research is to generate plausible hypotheses relevant to a person's goals – admittedly a weaker interpretation than Pearl's Causal Modeling Framework (Pearl, 2000, p. 43). This weaker interpretation is drawn from machine learning literature (Michalski, 1983; Quinlan, 1986; Holland, Holyoak, Nisbett, & Thagard, 1986), and serves as the basis for the noncompensatory policy capturing technique presented in this paper.

In Hammond and his colleagues' formulation of the lens model, inference from data to judgment policy is performed using linear regression (Hammond, Hamm, Grassia, & Pearson, 1987) – an analogous technique is needed for the noncompensatory policy capturing. A review of machine learning methods (Michalski, 1983; Quinlan, 1986; DeJong, Spears, & Gordon, 1993; Vafaie & DeJong, 1994), showed that genetic algorithms (GAs) tend to be more robust in concept learning applications as well as being better performers than other machine learning methods (Chen, Shankaranarayanan, She, & Iyer, 1998; Greene & Smith, 1993). Moreover, GAs have the added advantage that search is done on the encoding of the genetic strings, not the strings themselves (Goldberg, 1989) – as will be discussed later in this section. This search methodology plays an essential role in describing the degree of satisficing within a potential judgment strategy.

To infer noncompensatory judgment policies, a more specific definition of induction can be found in genetic algorithm literature. Holland et al. (1986) suggest that induction is a process of revisiting existing condition-action rule parameters and generating new rules based on knowledge about environmental variability (p. 22). Each rule is defined to represent a unit of knowledge, and collections of rules serve to represent internal states of the learning system (p. 15). Similarly, we define induction as a process of modifying a population of rule sets representing candidate judgment strategies. The rule sets are generated and modified on the basis of empirical data representing actual instances of human judgment—which we call exemplars. In

the following sections, we introduce and describe the technique we have developed for inducing noncompensatory judgment policies from exemplars, which we call Genetics-Based Policy Capturing (GBPC).

A. An Inductive Inference Model of Judgment

To be consistent with the findings from Gigerenzer and Goldstein (1996), the representation of each rule set in GBPC is of disjunctive normal form (DNF). To illustrate the form of a rule set, consider a conjunctive rule as a condition-action rule with N statements where:

$$IF (statement\ 1) \cap (statement\ 2) \cap \dots \cap (statement\ N) THEN (consequence\ statement)$$

A disjunctive rule is a condition-action rule with M statements where:

$$IF (statement\ I) \cup (statement\ II) \cup \dots \cup (statement\ M) THEN (consequence\ statement)$$

Each rule set in the population is represented as a disjunctive rule where each individual statement (e.g., *statement II*) is a conjunctive rule. DeJong, Spears, and Gordon (1993) demonstrated that a genetic algorithm (GA) was able to learn condition-action rules such as those above based on exemplars, and with little *a priori* knowledge of the exemplars themselves. Moreover, because GBPC rule sets are in DNF, the outcomes can potentially reflect not only fast and frugal heuristics, but also any logical strategy consisting of AND, OR, or NOT operators (Mendelson, 1997).

GBPC maintains a population of rule sets, where each rule set consists of a disjunction of conjunctive rules in DNF. Each rule within the rule set is a similarity template—or schema (Holland, 1975; Goldberg, 1989)—and is covered by the ternary alphabet $\{0,1,\#\}$ where “#” is a match-all character. The population of rule sets is trained on exemplars representing instances of human judgment. The instances consist not only of the human judgments themselves, but also cues available at the time of judgment. As in regression-based judgment modeling, the correct identification of cues that actually support human judgment is crucial to the success and utility of the modeling technique.

An example of a simple judgment domain will be used to illustrate the concepts underlying the induction approach and to clarify implementation details. Consider the case of a private pilot who is flying near a small airfield. The pilot sees four aircraft during the course of his flight, and makes judgments of their identity on the basis of two cues – speed and altitude – which we assume to be perceptually measured or encoded in a binary fashion, to simplify the discussion. The aircraft characteristics and corresponding pilot judgments are shown in Table 1.

Table 1. Aircraft Characteristics and Pilot Judgments for Sample Domain

SPEED	ALTITUDE	JUDGMENT
Fast	Low	racing aircraft
Fast	High	racing aircraft
Slow	High	transport aircraft
Slow	Low	transport aircraft

This judgment data will be used to demonstrate components of the inductive inference model and also the binary representation of candidate rule sets.

Consider a binary string representation for the judgments in the sample domain. Using the coding scheme where 1=fast, 0=slow, 1=high, 0=low, 1=racing aircraft, and 0=transport aircraft, each of the judgments in Table 1 can be converted into the four exemplars shown in Table 2.

Table 2. Exemplar Representation for Sample Domain Judgments

Exemplar No.	Characteristics Represented	Exemplar Representation
1	Fast, Low, racing aircraft	101
2	Fast, High, racing aircraft	111
3	Slow, High, transport aircraft	010
4	Slow, Low, transport aircraft	000

For example, Exemplar No. 1 represents the following operator judgment:

$$(Speed = Fast) \cap (Altitude = Low) \cap (Judgment = racing\ aircraft)$$

For simplicity's sake, the consequence statement is represented as part of the conjunctive statement. In addition, for the sake of illustration, consider the rule sets shown in Table 3 as the population used to learn the exemplars in the sample domain. Each rule set in the population represents data which genetic operators manipulate to form improved rule sets in future generations. For example, Rule Set No. 1 is represented by the following string: 1#10#0. The first three characters of the string (Speed=1, Altitude=#, Judgment=1) translates to the first rule:

$$(Speed = Fast) \cap (Altitude = anything) \cap (Judgment = racing\ aircraft)$$

Similarly, the next three characters of the string (Speed=0, Altitude=#, Judgment=0) translates to the second rule:

$$(Speed = Slow) \cap (Altitude = anything) \cap (Judgment = transport\ aircraft)$$

The rules combine in disjunctive normal form to create the following disjunctive rule set:

$$\begin{aligned} &(Speed = Fast) \cap (Altitude = anything) \cap (Judgment = racing\ aircraft) \\ &OR \\ &(Speed = Slow) \cap (Altitude = anything) \cap (Judgment = transport\ aircraft) \end{aligned}$$

Note that Rule Set No. 1 matches all the exemplars shown in Table 3.

Table 3. Sample Domain Rule Sets

Rule Set No.	Rule Set Representation
1	1#10#0
2	1#1
3	0#0
4	###
5	101111010000
6	11#100001000
7	000
8	001
9	010
10	011
11	100
12	101
13	110
14	111

To reflect bounded rationality, GBPC uses the Pittsburgh approach (Smith, 1983; DeJong, Spears, & Gordon, 1993) to learn. In the Pittsburgh approach, each rule set is a candidate judgment strategy and each strategy has a variable number of rules – hence strategies with a few simpler and effective rules reflect a satisficing mode of interaction. Learning consists of applying genetic operators in the order outlined by Goldberg (1989). That is, each learning cycle, also known as a generation, consists of: 1) fitness evaluation; 2) reproduction; 3) crossover; and 4) mutation. The multi-objective fitness evaluation process will be discussed later. Reproduction is achieved through use of a roulette wheel where rule set slots are apportioned based on fitness (for details see Goldberg, 1989). Mutation is implemented through random alteration of a bit in a rule set.

GBPC uses the variable-length 2-point crossover operator developed by DeJong and Spears (1990). The operator was shown to be effective by DeJong et al. (1993). This crossover operator selects a pair of rule sets, and then selects two positions within each rule set to exchange information. The positions are constrained only by the relative distance from the beginning and end of each rule set. For example, given Rule Set Nos. 5 and 1 and four randomly selected positions (indicated by |):

Rule Set No. 5: 10 | 11110100 | 00

Rule Set No. 1: 1# | 10 | #0

The resulting rule sets after applying the crossover operator are:

Rule Set No. 5: 10 | 10 | 00

Rule Set No. 1: 1# | 11110100 | #0

Through 2-point crossover, information between viable rule sets within GBPC are exchanged.

B. Fitness Evaluation

A central element of GBPC is the multi-objective fitness evaluation function. As mentioned earlier, the search in genetic algorithms is done on the encoding of the strings, and not the strings themselves. Hence, the quality of decisions (i.e., the form of the encoding) can be captured. In the traditional linear regression approach to judgment modeling, the best fitting linear-additive judgment rule is determined by least-squares. In moving to a noncompensatory approach to judgment modeling, we must define an alternative to least-squares for measuring the goodness or “fitness” of a rule set. In a subsequent section of this paper, we provide some evidence for the plausibility of this fitness evaluation measure, by showing that the rule sets induced by GBPC in a laboratory CIC simulated task were consistent with human judgment data.

We start by considering the fitness of a rule set as the ability to classify a set of exemplars in a manner consistent with satisficing behavior within bounded rationality. Therefore, a rule set should not only match a set of exemplars, but it should also resemble the types of noncompensatory judgment strategies performers typically use as heuristics in these tasks. This is done in GBPC through the use of a multi-objective function that evaluates fitness along three dimensions: completeness, specificity, and parsimony. The completeness dimension is based on work by DeJong et al. (1993), and is a measure of how well a rule set matches the entire set of exemplars (i.e., human judgments in a data set). The specificity dimension was first suggested by Holland et al. (1986), and is a measure of how specific a rule set is with respect to the number of wild cards it contains. Therefore, rule sets with less match-all (i.e., “#”) characters are classified as more specific. The parsimony dimension is a measure of the goodness of a rule set in terms of the necessity of each rule. Hence, in a parsimonious rule set, there are no unnecessary rules. The ideal rule set, therefore, will match all operator judgments, will be maximally specific, and maximally parsimonious. The mathematical formulation of each dimension will be discussed in the following section.

C. Mathematical Development of Fitness Dimensions

Definition 1.1: An exemplar matrix, \mathbf{E} , consisting of a set of binary variable vectors, called exemplars, whose range is the set $\{0,1\}$. Each exemplar within \mathbf{E} is represented as $\mathbf{e}_{i\bullet}$, for $i = 1, \dots, m$, where m is the total number of exemplars. Each binary variable within $\mathbf{e}_{i\bullet}$ is represented as e_{ij} for $j = 1, \dots, n$, where n is exemplar length. Thus, \mathbf{E} is a m by n matrix in the form:

$$\mathbf{E} = \begin{bmatrix} e_{1,1} & e_{1,2} & \cdots & e_{1,n} \\ e_{2,1} & e_{2,2} & \cdots & e_{2,n} \\ \vdots & \vdots & & \vdots \\ e_{m,1} & e_{m,2} & \cdots & e_{m,n} \end{bmatrix} \quad (2)$$

Definition 1.2: A rule set matrix is a matrix, \mathbf{S} , consisting of a set of ternary variable vectors, called a rule set, whose range is the set $\{0,1,\#\}$. Each rule within \mathbf{S} is represented as $\mathbf{s}_{k\bullet}$ for $k = 1, \dots, p$, where p is the number of rules in the rule set. Each ternary variable within $\mathbf{s}_{k\bullet}$ is represented as s_{kj} , for $j = 1, \dots, n$ where n is the rule length.

Therefore, \mathbf{S} is a p by n matrix in the form:

$$\mathbf{S} = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & & \vdots \\ s_{p,1} & s_{p,2} & \cdots & s_{p,n} \end{bmatrix} \quad (3)$$

Matching an exemplar with a rule set requires an indicator function. Therefore, given that $x \in \mathbf{E}$ and $y \in \mathbf{S}$, an indicator function is defined as I_A where $A = \{x, \#\}$, so that,

$$I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{if } y \notin A \end{cases} \quad (4)$$

The results of applying I_A to compare a rule set, \mathbf{S} , with the i th exemplar $e_{i\bullet}$, is shown as a “matching matrix”, \mathbf{M}_i , for exemplar i where

$$\mathbf{M}_i = \begin{bmatrix} I_{\{e_{i,1}, \#\}}(s_{1,1}) & I_{\{e_{i,2}, \#\}}(s_{1,2}) & \cdots & I_{\{e_{i,n}, \#\}}(s_{1,n}) \\ I_{\{e_{i,1}, \#\}}(s_{2,1}) & I_{\{e_{i,2}, \#\}}(s_{2,2}) & \cdots & I_{\{e_{i,n}, \#\}}(s_{2,n}) \\ \vdots & \vdots & & \vdots \\ I_{\{e_{i,1}, \#\}}(s_{p,1}) & I_{\{e_{i,2}, \#\}}(s_{p,2}) & \cdots & I_{\{e_{i,n}, \#\}}(s_{p,n}) \end{bmatrix} \quad (5)$$

Each row of the matching matrix represents how well an exemplar $e_{i\bullet}$ matches a particular rule, $s_{k\bullet}$, within the rule set. To simplify, rewrite \mathbf{M}_i so that each element is represented as a binary variable, $m_{i,k,j}$, such that,

$$m_{i,k,j} = I_{\{e_{i,j}, \#\}}(s_{k,j}) \quad (6)$$

Before elements of the matching matrix can be algebraically manipulated, one first needs to show that the matrix is a lattice under conjunction and disjunction.

Theorem 1.1: A matching matrix, \mathbf{M}_i , is a boolean lattice under disjunction, \cup , and conjunction, \cap .

Proof: First, it is seen that \mathbf{M}_i is ordered by the relation \leq so that, for each pair a, b of binary variables in \mathbf{M}_i , $a R b \Leftrightarrow a \leq b$. It follows that \mathbf{M}_i is an ordered set. Second, the supremum and infimum of each pair of binary variables can be readily determined as either 0 or 1. Note that the supremum and infimum are, effectively, the disjunctive and conjunctive operators, respectively. Thus, the proof is complete.

Given a lattice, \mathbf{a} where $\mathbf{a} = \{a_1, a_2, \dots, a_n\}$, the disjunct, $a_1 \cup a_2 \cup \dots \cup a_n$, and the conjunct, $a_1 \cap a_2 \cap \dots \cap a_n$, are denoted by $\bigcup_{i=1}^n a_i$ and $\bigcap_{i=1}^n a_i$, respectively. Thus, by applying the

disjunct operator on elements of the matching matrix, we define that an exemplar is matched to a rule if and only if:

$$\bigcap_{j=1}^n m_{i,k,j} = 1 \quad (7)$$

That is, for any exemplar i , and any rule k , both having length n , a vector-wise match exists if and only if each exemplar value matches the corresponding rule value. Thus, a matching function, f , between a rule set, \mathcal{S} , and an exemplar $e_{i\bullet}$, can be formulated as,

$$f(M_i) = \bigcup_{k=1}^p \left[\bigcap_{j=1}^n m_{i,k,j} \right] \quad (8)$$

For p rules in the rule set, each with length n . A match, therefore, between an exemplar $e_{i\bullet}$ and a rule set exists if and only if $f(M_i) = 1$.

Definition 1.3: A rule set is said to be complete if it is able to match all the exemplars in the exemplar set. A scaled function to indicate rule set completeness, c , follows:

$$c(M_1, M_2, \dots, M_r) = \frac{\sum_{i=1}^r f(M_i)}{r} \quad (9)$$

For r exemplars. Thus, $0 \leq c \leq 1$. Completeness values for all rule sets shown in Table 3 are listed in Table 4.

Table 4. Sample Domain Rule Set Completeness Values. Exemplars (see Table 2 consist of {101,111,010,000})

Rule Set No.	Rule Set String	c
1	1#10#0	1
2	1#1	0.5
3	0#0	0.5
4	###	1
5	101111010000	1
6	11#100001000	0.5
7	000	0.25
8	001	0
9	010	0.25
10	011	0
11	100	0
12	101	0.25
13	110	0
14	111	0.25

Therefore, c discriminates between rule sets not matching any exemplars (Rule Set Nos. 8, 10, 11, and 13), rule sets matching some exemplars (Rule Set Nos. 2, 3, 6, 7, 9, 12, and 14),

and rule sets matching all the exemplars (Rule Set Nos. 1, 4, and 5). Although three rule sets are able to match all exemplars, the usefulness of each rule set as a judgment strategy varies greatly. Rule Set No. 4 represents an over-generalized strategy (i.e., if anything do anything). Rule Set No. 5 represents a strategy that relies on memorization of all possible outcomes, which may be theoretically possible, though practically prohibitive in a complex, dynamic environment. Rule set No. 1 represents a simplification strategy consistent with the findings in (Klein, 1999; Rasmussen, 1983; Rouse, 1983; Kirlik, Walker, Fisk, & Nagel, 1996).

Thus, although the completeness function is able to measure the degree to which a rule set matches an exemplar set, a well-matched rule set does not necessarily represent a cognitively plausible judgment strategy. Two other fitness dimensions will now be introduced in an attempt to improve the capability of the fitness function to better achieve psychological plausibility. The specificity dimension addresses the task of eliminating rules within a rule set that are over-generalized (e.g. Rule Set No. 4).

Definition 1.4: A rule set is fully specified if there are no match-all characters in the rule set. A scaled function to show rule specificity, t , follows:

$$t(S) = \frac{\sum_{k=1}^p \sum_{j=1}^n I_{\{0,1\}}(s_{k,j})}{(p * n)} \quad (10)$$

For p rules of length n each. Thus, $0 \leq t \leq 1$. Specificity values for all rule sets shown in Table 3 are listed in Table 5.

Table 5. Sample Domain Rule Set Specificity Values

Rule Set No.	Rule Set String	t	$c * t$
1	1#10#0	0.6667	0.6667
2	1#1	0.6667	0.3334
3	0#0	0.6667	0.3334
4	####	0	0
5	101111010000	1	1
6	11#100001000	0.9167	0.4584
7	000	1	0.25
8	001	1	0
9	010	1	0.25
10	011	1	0
11	100	1	0
12	101	1	0.25
13	110	1	0
14	111	1	0.25

As a complement dimension to c , t discriminates between rule sets not containing match-all characters (Rule Set Nos. 5, 7-14), and rules sets that do (Rule Set Nos. 1-4, 6). Table 5 also shows a combined completeness/specificity value (in the form of $c*t$). An examination of $c*t$ shows that the fitness of the over-generalized rule (Rule Set No. 4) has been reduced in value.

Furthermore, the two highest $c*t$ rule sets (Nos. 1 and 5) continue to support possible decision strategies as measured by completeness. However, Rule Set No. 6 presents another difficulty that must be overcome. Although half of the rules in Rule Set No. 6 match exemplars in Table 2, the other rules do not. Nevertheless, the two useless rules contribute to the overall specificity value of the rule set. Therefore, the final fitness dimension, parsimony, will now be introduced to eliminate useless rules from the rule set.

Definition 1.5: A rule set is said to be parsimonious if each rule within the rule set matches at least one exemplar in the exemplar set. A scaled function to indicate rule set parsimony, p , follows:

$$p(M_1, M_2, \dots, M_r) = \frac{\sum_{k=1}^q \left[\bigcup_{i=1}^r \left(\bigcap_{j=1}^n m_{i,k,j} \right) \right]}{q} \quad (11)$$

For r exemplars and q rules of length n each. Thus, $0 \leq p \leq 1$. Parsimony values for all rule sets shown in Table 3 are listed in Table 6.

Table 6. Sample Domain Rule Set Parsimony Values

Rule Set No.	Rule Set String	p	$c * t * p$
1	1#10#0	1	0.6667
2	1#1	1	0.33335
3	0#0	1	0.33335
4	###	1	0
5	101111010000	1	1
6	11#100001000	0.5	0.22918
7	000	1	0.25
8	001	0	0
9	010	1	0.25
10	011	0	0
11	100	0	0
12	101	1	0.25
13	110	0	0
14	111	1	0.25

The third fitness dimension, p , provides a means of discriminating between rule sets that are wholly useful (i.e., each rule matches at least one exemplar) and those that are not. Table 6 also shows a combined completeness/specificity/specificity value (in the form of $c*t*p$). Notice that the fitness value of Rule Set No. 6 has been reduced to correspond to the usefulness of each rule within the rule set. Thus, an examination of Table 6 shows that the two highest $c*t*p$ rule sets (Nos. 1 and 5) represent viable decision strategies to judge the identity of an aircraft based on its speed or altitude attributes as outlined in the sample domain.

Interestingly, the two rule sets (Nos. 1 and 5) represent disparate decision strategies. Rule Set No. 1 represents a simplification policy that identifies aircraft based strictly on speed while No. 5 generates a comprehensive set of specific rules to describe each exemplar exactly. Because

the multi-objective fitness function in GBPC does not prescribe the number of rules to generate, the maximal number of rules employed by an operator should be empirically determined. Regardless, GBPC provides candidate rule sets that represent plausible strategies based on concepts of completeness, parsimony, and specificity. A global function using all three fitness dimensions will now be discussed.

Definition 1.6: A global fitness function, \mathbf{g} , combines all three fitness dimensions, and is defined as,

$$g = c^2 * t^2 * p^2 \quad (12)$$

Thus, \mathbf{g} provides a non-linear differential reward system for rule sets within the population. The global maximum of $\mathbf{g}=1$ is achieved when all exemplars are fully contained in a rule set (e.g., Rule Set No. 5). While \mathbf{g} was initially selected for its simplicity, further studies are underway to explore alternative formulations (Rothrock & Repperger, in review). For the present purposes, it is important to note that the particular formulation for \mathbf{g} used here, in which the values of each of the contributing terms are squared prior to summation, was selected for computational rather than psychological reasons. This choice has implications for how to fairly compare the scalar measure of noncompensatory model fitness as represented by \mathbf{g} and analogous scalar measures of regression model fitness, such as multiple correlation, as will be seen in the following section. We next describe an empirical evaluation of GBPC for inferring noncompensatory judgment rules in a dynamic task.

V. EMPIRICAL EVALUATION

The inductive inference model was applied to human judgment data collected in a dynamic laboratory simulation of a U.S. Navy combat information center (CIC). Detailed information on the simulation and experimentation can be found in (Rothrock, 1995; Hodge, 1997). The simulation required participants to perform the tasks of an anti-air warfare coordinator (AAWC), responsible for making judgments about the identity of initially unknown vehicles (or “tracks”) entering his geographic area of responsibility. GBPC was used to infer the possible heuristic strategies used by AAWC participants to make these track identification judgments.

Participants consisted of university students who were initially briefed on the role of an AAWC operator, functions of the computer interface, and geopolitical context of the simulation. Participants were given maps and profiles of friendly and hostile aircraft in the area to study and, later, during the scenario runs. Subjects were also briefed on the relative diagnosticity of each type of cue. For example, subjects were told that visual identification is veridical. They were then trained on 15 30-minute scenarios of comparable difficulty. In the training scenarios, post-scenario feedback was provided to each participant in terms of correct assessments and incorrect actions. Participants then ran three additional 30-minute scenarios during which data for GBPC was collected. The number of identification judgments per scenario ranged from 15-34.

Participants were provided with a radar display and a suite of controls for obtaining additional information about tracks in the vicinity of their ship. The following types of information, or judgment cues, were available: a) Identification Friend or Foe (IFF) status; b)

electronic sensor emissions (friendly or hostile sensor onboard); c) visual identification by combat air patrol (CAP); d) range; e) altitude; f) speed; g) course; h) bearing; i) location of civilian airports and air corridors; and j) legitimate commercial aircraft flight numbers; and k) designation of hostile and friendly countries. The participant's goal was to use the available information to identify initially unknown tracks as either friendly, assumed friendly, hostile, or assumed hostile. For the simulation, we intentionally made some of this information more diagnostic than others. For example, visual identifications provided by CAP were perfectly diagnostic, and electronic sensor emissions were highly diagnostic. IFF information, however, was much less reliable. Our experimental purposes did not require that we mimic the actual, relative reliability of these information sources in the operational naval environment, as we were not attempting to actually train naval personnel using this simulation. Naturally, however, a training simulation should mimic the reliability of information sources in the target context.

A. Modeling Approach

To evaluate GBPC, both the regression-based Lens Model technique and the GBPC technique were fit to empirical data from the highest (A) and lowest (B) performing participants in the track identification task in a final experimental session (i.e., after judgment strategies had presumably stabilized). Participant A judged the identity of 20 out of 24 possible tracks, made no errors, and did not judge any track multiple times. Participant B judged only 14 out of the 24 tracks, made four errors, and judged five tracks twice. In every instance where a track was judged twice, the second judgment was made when visual identification became available. The goal in analyzing the behavior of these two participants using both modeling approaches was to compare the results of the two methods to see if they revealed similarities or differences in representing judgment strategies as well as in explaining the performance differences between the participants. The bottom line in this investigation was to try to determine why participant A performed this dynamic judgment task more successfully than participant B, and also to exploit the opportunity provided by a common data set to compare the GBPC approach with the much more established Lens Modeling approach.

Due to the large number of information sources available in the laboratory task, we first divided these sources into two categories: *active* and *passive*. Active information sources required the operator to make queries about a track. Active sources included queries of IFF, electronic sensor emissions, and requests for visual identifications of a track, obtained by sending CAP resources to fly to a track's location and make a report, if possible, to the AAWC. All other information sources were considered to be passive, since they did not have to be actively requested, but were instead continuously available from the radar display (e.g., track location, bearing, speed, altitude, etc.). The first stage of modeling focused solely on the performers' use of active information. We parsed the data in this fashion due to the limited number of human judgments available (a maximum of 24), which meant that we would have to focus on a relatively small set of cues or information sources in order to give us the chance to obtain reliable fits for either GBPC and Lens Modeling. For GBPC, there is a potential for combinatorial explosion when representing the laboratory task in the binary format required by GBPC. For Lens Modeling, there is a need for a relatively high ratio of the number of judgments to the number of cues in order to construct a reliable model.

We must note that by restricting the cue data set in this way we did not expect to create complete accounts of our performers' judgment strategies. However, focusing on just the set of active information not only enabled the comparison of GBPC and Lens Model results, but also helped address the question of whether the difference between the high and low scoring participants could have been due to their policies for actively searching for judgment cues, and how they may have used these cues. As will be described in a following section, however, we did conduct a second stage of GBPC based on the use of both active and passive sources of information. While that second modeling effort is not the primary focus of this article, selected findings from that second, more comprehensive GBPC analysis will be presented – in particular, those that bear on diagnosing the possible task-simplification heuristics underlying participant B's erroneous judgments. A complete account of both stages of GBPC is provided in (Rothrock, 1995).

B. Coding the Active Information Data Set to Support GBPC and Lens Model Analysis

For modeling the use of active information, cues and operator judgments were encoded in GBPC as a 10-bit string. The meaning of each string position is shown in Table 7. Note that the first six bits represent actions taken to seek judgment cues (and in some cases, the information gained as a result of these actions), while the last 4 bits represent the four possible AAWC identification judgments themselves. Note that representation provided in Table 7 is hardly the most efficient binary coding from an information theoretic perspective. However, alternative, more efficient codings may limit the representational flexibility of the model, and thus limit its ability to induce rule sets covering the entire range of exemplars in a data set.

Table 7. AAWC String Bit Position and Representation

Bit	Representation	1	0
#1	IFF queried	yes	no
#2	Friendly emitter response	yes	no
#3	Hostile emitter response	yes	no
#4	Negative emitter response	yes	no
#5	Friendly visual sighting	yes	no
#6	Hostile visual sighting	yes	no
#7	Friendly AAWC judgment	yes	no
#8	Assumed friendly AAWC judgment	yes	no
#9	Assumed hostile AAWC judgment	yes	no
#10	Hostile AAWC judgment	yes	no

To support Lens Modeling of these same data, we used a coding approach very similar to that presented in Table 7, but modified in such a way as to be consistent with the requirements of linear regression modeling, and similar to the manner in which we coded cue and judgment data in our previous Lens Model analysis of a different data set collected in the same laboratory task (Bisantz, Kirlik, Gay, Phipps, Walker, & Fisk, 2000). In particular, friendly tracks (both judgments and criterion) were coded with the value of 1.0, and hostile tracks with the value of –

1.0. IFF, emission, and sighting cues were coded as either 1.0, -1.0, or 0.0, depending on whether the associated cue provided evidence of a friendly track, a hostile track, or neutral (or no) evidence on track identity, respectively. It should be noted that the data set modeled in this paper, while distinct from that used in the Bisantz et al. (2000) paper, only differed in the fact that it reflected behavior from a different set of performers collected at a different time. Both data sets came from control groups from a series of experiments evaluating a variety of training interventions for the CIC context.

C. Lens Model Analysis of the Use of Active Information

Recall that participant A made 20 identification judgments, all correct, while participant B made 19 identification judgments, with four errors. Lens Models of both participants A's and B's strategies for the use of active information were reliably created (for A, $R^2 = 0.851$, $R^2(\text{adj}) = 0.834$, $F(2,19) = 48.74$, $p < 0.001$; for B, $R^2 = 0.602$, $R^2(\text{adj}) = 0.522$, $F(2,19) = 7.57$, $p < 0.01$). In keeping with the manner in which Lens Model results were graphically presented in the previous study in the same task (see Fig. 8 in Bisantz et al., 2000), Figure 2 provides a comparison of performers A and B in terms of the Lens Model measures of cognitive control (or consistency), environmental predictability, achievement, linear knowledge, and unmodeled knowledge.

There are many notable findings in Figure 2. First, consider the findings regarding participant A. Naturally, Lens Modeling revealed a perfect (unity) achievement measure for A, as he made no judgment errors. The best fitting model for A indicated heavy reliance on both the sensor emission cue ($\beta = 1.00$, $p < .001$) and the visual identification cue ($\beta = .940$, $p < .01$). Since A never queried IFF, there was no variance in this cue value so it was therefore not included in A's model. In addition, A demonstrated a degree of cognitive control (.922) that was exactly equal to environmental predictability (.922), a necessary result because of the fact that since A scored perfectly, his judgment model and the environmental regression models were identical. In addition, A demonstrated a perfect (unity) degree of linear knowledge for exactly the same reason (the β weights in both models were identical). All these findings would lead one to suspect that the compensatory judgement model for A provided a very good description of his performance, except for the striking value of C (also unity), indicating a highly significant degree of unmodeled knowledge profitably used by this participant. In summary, the general interpretation of A's behavior invited by this Lens Model analysis is that he was well adapted to weighting and combining the linear cue-criterion relationships, and used some additional knowledge about some amount additional non-linear relations between these cues and the criterion to overcome the less-than-perfect linear predictability of his environment, and his less than perfect cognitive control in executing his judgment strategy.

In contrast, now consider the interpretation of participant B's performance based on these Lens Model results. B's achievement of .567 represents a considerably lower degree of judgment performance, due to errors committed. The best fitting model for B indicated heavy reliance on both the sensor emission cue ($\beta = 0.958$, $p < .01$) and the visual identification cue ($\beta = 1.266$, $p < .05$). Although there was variance in the IFF cue for B due to his occasional use of IFF queries, no reliable linear weighting of this cue was found, so it was not included in the final model of participant B. The explanation provided by these results suggests that a lack of

complete cognitive control (.783), environmental predictability (.776), and linear knowledge (.894) all contributed to B's relatively modest achievement in this task.

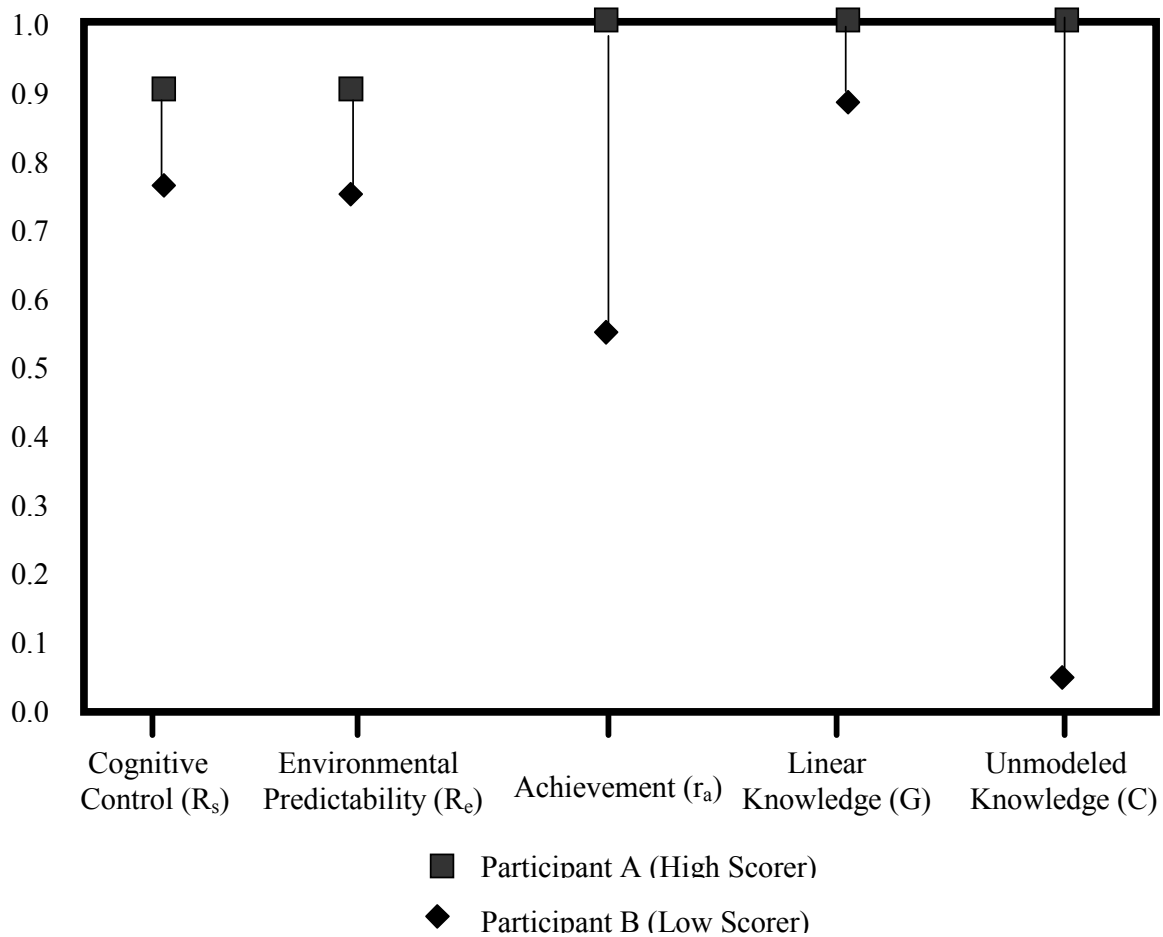


Figure 2. Lens Model Results for High & Low Scoring Participants Using Active Information

Figure 2 also supports a comparative assessment of A's and B's judgment strategies. While A displayed perfect use of unmodeled (presumably, non-linear) knowledge, B displayed the use of no such knowledge. We note that a large difference in reliance on unmodeled knowledge was also found in a Lens Model analysis of another data set collected in the same experimental context (see Fig. 8 in Bisantz et al., 2000), but those authors did not offer an explanation of why their high and low performers may have differed in this respect. Finally, note the different values for environmental predictability for A and B. This analysis suggests that although A and B were performing the "same" task, by using a more effective strategy for information search (recall these models were created on the basis of actively sought information), A was able to essentially perform in a more predictable, proximal environment than participant B. This finding is important, as it suggests that one component of judgment skill

in dynamic, interactive tasks is the use of an adaptive strategy for actively searching for diagnostic information.

In summary, it is clear that a Lens Model analysis of these two participants has provided some useful information, although, especially in the case of participant A, his perfect use of unmodeled knowledge should raise some suspicions about whether a fully compensatory description of his judgment strategy is faithful to the strategy he actually used.

D. GBPC Analysis of the Use of Active Information

When GBPC was applied to data from performers A and B and allowed to learn, GBPC produced a rule set for A with an overall fitness value, or $g = 0.5625$, and a rule set for B with $g = 0.3600$. The lack of fit (the difference between g and unity) for both operator models was due solely to the *specificity* dimension by which fitness was evaluated. GBPC inferred rule sets for performers A and B that were both fully complete (i.e., covered all judgment instances), and also fully parsimonious (i.e., contained no unnecessary rules). The rule set for A achieved a specificity value of 0.7500 (and thus an overall fitness g value of $\{1.0000^2 * 1.0000^2 * 0.7500^2\} = 0.5625$), and a rule set for B with a specificity value of 0.600 (and thus a g value of $\{1.0000^2 * 1.0000^2 * 0.6000^2\} = 0.3600$). A lack of specificity suggests that both operators used abstract heuristics to generalize their strategies (that is, that some of the rules in their final rule sets referred only to a subset of the three (IFF, sensor, and visual identification) cues. Finally, in regard to the numerical measures of fit obtained by GBPC, recall that the equation whereby each of the three contributing fitness measures are squared prior to summation was done purely for mathematical convenience, and not for any psychological reason. Thus, it may be just as plausible to assess the GBPC fits without squaring, resulting in a fit for participant A of .750 and a fit for participant B of .600).

Given that this first stage of modeling inferred judgment strategies on the basis of a highly restricted set of cues (IFF, sensor emissions, and visual identifications), we were surprised to achieve these high (unity) fitness values on the *completeness* fitness dimension. Recall that A made a total of 20 judgments, and GBPC found a disjunctive normal form (DNF) representation of this operator's strategy as a disjunction of seven conjunctive rules, with one of these rules covering only one judgment instance. The remaining six rules in this rule set covered between two and eight instances. Operator B made a total of 19 judgments, and GBPC found a DNF representation of this operator's strategy as a disjunctive collection of 11 rules, with one of these rules covering only one judgment instance. The remaining 10 rules for Operator B covered between two and eight instances.

Analysis of the rule sets with maximum g values (i.e., the winning rule sets) revealed some interesting findings regarding the use of active information. Operator A's winning rule set indicated a reliance on querying electronic sensor emissions and visual identifications to make track identification judgments. This finding is consistent with the Lens Model analysis, although that analysis implies a compensatory rather than a noncompensatory reliance on these cues. For example, the following two rules from the winning rule set, which covered eight and nine of A's judgments respectively, indicate reliance on these highly diagnostic sources of information to make "hostile" and "assumed friendly" judgments.

$(\text{friendly emitter response} = \text{no}) \cap (\text{emitter response} = \text{not negative}) \cap (\text{AAWC judgment} = \text{hostile})$
 OR
 $(\text{emitter response} = \text{not negative}) \cap (\text{hostile visual ID} = \text{no}) \cap (\text{AAWC judgment} = \text{assumed friendly})$

Recall that GBPC represents rule sets as disjunctions of conjunctive rules. The first conjunctive rule above states that if a sensor emission is queried, and the response is neither friendly nor negative, then judge the associated track to be hostile. For all the tracks in the experimental task, correct sensor assessment yields correct identifications. Hence, the rule is diagnostic of the true identity of the tracks.

Now consider the second conjunctive rule above. This rule states that if a sensor is queried, and some emission (either friendly or hostile) is detected, and CAP resources do not provide a visual identification of the track as hostile, then assume the track to be friendly. This rule has two interpretations. In the first case, assume that the sensor emission is friendly. In this case, the track should be judged as "assumed friendly" given that visual identification provided by CAP does not indicate otherwise (i.e., does not indicate that the track is hostile). This rule thus represents a reliance on highly diagnostic emission information unless the (even more diagnostic) visual identification conflicts with emission information, and is thus fully consistent with the relative diagnosticity of these two cues in this task environment.

Now, consider the second interpretation of this rule, namely, that the emission response is hostile. In this case, the track should instead be assumed to be friendly if CAP provides a visual identification to override the judgment that would be made on the emission report alone (e.g., the first of the two disjunctive rules above). This second interpretation of this rule can also be seen as a refinement of the first rule, as it indicates that information gained by visual identification, if available, should override any judgments made solely on the basis of electronic sensor emissions. In the experimental task, sensor emissions were highly diagnostic, as mentioned above, but visual identifications were 100% diagnostic. Thus Operator A's second rule reflects an adaptive refinement of his first rule to those cases where CAP resources provide conflicting visual identification information. As compared to the Lens Model analysis that indicated that A made highly significant (but unexplained) use of unmodeled knowledge, we believe that this noncompensatory, rule-based description of A's judgment strategy may be the more plausible one, although this restricted data set is clearly insufficient to establish this conclusion with certainty.

Finally, the rule set for A also indicated a generally adaptive lack of reliance on relatively unreliable IFF information in making judgments: the only two rules in his rule set that matched judgments where IFF had been queried also relied upon sensor information, visual identification information, or both sources, to complement any information obtained from IFF.

When GBPC was applied to participant B's use of active information, on the other hand, it was much more difficult to infer a coherent and efficient judgment strategy on the basis of active information alone. At a general level, however, B did not appear to make effective use of the most highly diagnostic types of active information: electronic sensor emissions and visual identifications. In fact, one rule, covering six instances of B's judgments, indicated a reliance solely on unreliable IFF information, with no accompanying reliance on sensor or visual information to supplement IFF as an information source:

$(\text{IFF query status} = \text{anything}) \cap (\text{AAWC judgment} = \text{assumed friendly})$

We emphasize, however, that GBPC was less successful in inferring an efficient judgment strategy for participant B than for A, solely on the basis of their use of active information sources. Other than a general over-reliance on relatively unreliable IFF information, the inferred strategy for B was not particularly enlightening as to the possible reasons underlying his four judgment errors. Once again, these results are consistent with the Lens Model analysis of B, as, in direct opposition to A, B's demonstrated lack of any reliance on linearly unmodeled knowledge suggests that moving toward a noncompensatory formulation of B's strategy would likely have limited success.

E. Modeling the Use of Passive Information: Inferring Error Tendencies

To gain additional insight into the differences between the judgment strategies that may have been used by performers A and B, a second stage of GBPC was performed that also included the seven dimensions of passive information discussed previously. This was a significantly more elaborate and computationally-intensive exercise, due to the need to represent the seven passive information sources in a binary format that was hopefully consistent with how the operators perceived and encoded information obtained from the radar display. We eventually settled on a 40-bit representation of the task environment for this second stage of modeling.

The first 8 bits in the 40-bit string represented the 8 different radar ranges (e.g., 8 nm, 16 nm, etc.) which could be selected by the operator. The ninth bit represented whether a track was emitting electronic sensor information. Bits 10 through 13 represented a track's altitude, put into equivalence classes that were somewhat relevant to a track's identity in this environment (e.g., less than 5000 feet, between 5000 and 18,000 feet, etc.). Bits 14 through 17 represented a track's speed in a similar format. Bits 18 through 21 represented a track's course as one of four compass quadrants. Bits 22 through 25 similarly represented a track's bearing. Bits 26 through 33 represented a track's range as a member of one of eight, task-relevant, equivalence classes. Bits 34, 35, and 36 represented whether IFF, electronic sensor emission, and visual identification information for a track had been sought by the operator. Finally, the last four bits represented the operator's judgment, in the same manner as used in the first modeling stage. Because of the large numbers of information sources used in this stage of modeling, statistical power was lacking to conduct analogous Lens Model analyses of these data for comparison purposes.

A detailed discussion of the results of this second stage of GBPC modeling can be found in (Rothrock, 1995). Here, the focus is solely on an analysis of the differences between the ways in which A and B made judgments about the identity of three particular tracks. Each of these three tracks was a hostile helicopter, correctly identified as hostile by A, but incorrectly identified by B as "assumed friendly." These three misidentifications accounted for three of B's four judgment errors. Based on GBPC results from this second modeling stage, rule sets were found suggesting that A correctly identified all three of these helicopters by sending CAP resources to obtain visual identifications. On the other hand, the rules which covered B's judgments about these helicopters indicated that no active information sources were sought for two of these helicopters, and that the third was queried only by relatively unreliable IFF. Additionally, the rule covering these three helicopter judgments for B contained the following information:

$$(altitude \text{ not } \{18,000 < a < 40,000\}) \cap (altitude \text{ not } \{a > 40,000\}) \cap (speed < 200) \cap \\ (range \text{ not } \{r < 10\}) \cap (range \text{ not } \{40 < r < 50\}) \cap (range \text{ not } \{r > 150\}) \cap (assume \text{ friendly})$$

Of particular interest here are the track conditions described in this rule (speed less than 200 kts, altitude < 18,000 feet). This information was available from the radar display. These track conditions generally reflected the radar signature of commercial airliners taking off from airports in our simulation. All tracks in our scenarios with this signature were indeed airliners except for the three hostile helicopters misidentified by B. Recall that A did not solely use radar information to identify these tracks as hostile, relying instead upon actively sought, visual identification. Although one cannot be sure that the rule described above actually accounted for B's misidentification of these helicopters as "assumed friendly," this case does provide an example suggesting how inferential modeling might provide hypotheses about the nature of the task-simplification heuristics operators might employ, and how information gained from inferential modeling might provide an important source of feedback for training.

VI. CONCLUSIONS

A. Summary and Implications

Performers in time-stressed, information-rich environments develop heuristic, task-simplification strategies for coping with the time-pressure and often severe information processing demands of judgment and decision making tasks. Judgment strategies in these environments may have a noncompensatory nature, which may be adaptive to the time-stressed nature of these tasks, since such heuristics typically make lower demands for information search and integration than do corresponding, linear-additive, compensatory strategies. As a result, linear regression may be inappropriate for inferring the judgment strategies used by operators in time-stressed environments, assuming as it does that judgment strategies can be usefully described by compensatory, linear-additive rules.

An alternative approach for inferring judgment strategies from behavioral data has been presented that does not rely on the compensatory assumptions underlying linear regression. The technique, Genetics-Based Policy Capturing (GBPC), infers noncompensatory judgment strategies under the assumption that these strategies can be described as a disjunctive collection of conjunctive rules. The fitness measure embodied in GBPC evaluates candidate rule sets on three dimensions: a) *completeness* (the inferred rule base is consistent with all operator judgments); b) *specificity* (the rule base is maximally concrete); and c) *parsimony* (the rule base contains no unnecessary rules).

The inferential approach was illustrated using behavioral data from the highest and lowest performing operators of a laboratory simulation of a combat information center (CIC) task. In this application, the GBPC inferred individually valid, yet contrasting rule bases for these two operators. Additionally, the two inferred rule bases were consistent with these operators' patterns of both correct and incorrect judgments. Also, it was shown that the GBPC results provided a useful complement to a Lens Model representation of the same data. In some cases, we suggest that the GBPC results may have even provided a superior representation of judgment, for example, in explaining the highly significant reliance on (linearly) unmodeled knowledge demonstrated by the highest scoring participant.

GBPC holds promise for the design of advanced training technologies that use individual performance histories to target feedback toward eliminating any potential misconceptions or oversimplifications a trainee's behavior might reflect. One can imagine using both Lens Model analysis and GBPC analysis to capture trainee data in real time, infer judgment strategies as enough data on trainee behavior became available, and then make these strategies explicit to a human trainer or the trainee himself or herself as a form of feedback augmentation. As a knowledge engineering tool, the technique could be used to identify the judgment strategies used by expert performers in dynamic, time-stressed environments, when provided with a data set of expert judgments and the task conditions in which these judgments were made.

B. Towards a Noncompensatory Formulation of the Lens Model

Thus far, GBPC has been developed solely as a policy capturing technique (Cooksey, 1996, p. 57) to characterize a performer's cue utilization strategy. While the authors are encouraged by our results to date, we realize that merely fitting a sample data set does not validate a model. One of our future goals, therefore, is to cross-validate (Kohavi, 1995) the model through additional experimentation to evaluate the usefulness of the model in the context of a broader data set, and the use of held-out data sets that were not used in the model fitting process. Most importantly, however, we are also currently investigating the use of the GBPC to describe not only the human performer, but the task environment as well, in the spirit of the compensatory formulation of the Lens Model.

Thus, the next step in our research is to apply the GBPC technique to modeling ecologies with various types of cue-criterion or means-ends structure, and compare the resulting models with regression-based, compensatory models of the same environments. As discussed in Part III of this paper, it is well known that in specified conditions linear-additive models can effectively mimic the behavior of truly noncompensatory strategies to various degrees. With both compensatory and noncompensatory inferential techniques available for describing ecological structure, we should be in a much better position to analyze, describe, compare, and contrast the conditions under which both rule-based strategies as well as linear-additive strategies both succeed and fail as a function of the cue-criterion or means-ends structure of the task ecology. Naturally, we expect to accompany this analytical exercise with an experimental program using human judges to gain an understanding of the degree to which humans are sensitive to these factors and tradeoffs.

Ultimately, our overall goal is to provide a set of techniques for analyzing and assessing the adaptivity of rule-based judgment strategies with the same level of precision and formality that current versions of the Lens Model provide to support the analysis of linear-additive strategies. Since the Lens Model Equation (LME) is basically a decomposition of the correlation coefficient measuring task achievement, the same equation can be used to evaluate the output of combined GBPC modeling of both the human and environmental components of a judgment system. Such a combined model could be termed a Genetics Based Lens Model, or GBLM.

With such a model in hand, it promises to be quite interesting to model human performance in both linear-additive ecologies and noncompensatory ecologies with both regression-based and GBLM techniques, and decompose the resulting correlations using the LME. We expect this to prove to be a valuable exercise, as it may require a reinterpretation of

the traditional psychological meanings of the LME parameters. For example, when linear regression is used as the basis for traditional Lens Modeling, the second term in the LME provides a measure of non-linear, noncompensatory, or otherwise “unmodeled knowledge.” In contrast, if GBLM were to be applied to the same data set, it may be that much of the correlation between judgments and the criterion described as non-linear or unmodeled knowledge by traditional methods may be transformed into modeled knowledge, and thus reflected in the first, rather than the second, term in the LME. Now, however, the degree of correlation reflected in this first term is no longer “linear knowledge,” as it would be called in standard Lens Modeling, but instead either “rule-based” or noncompensatory knowledge in the GBLM approach.

Investigations such as these would thus raise the question of the appropriate psychological interpretation of the second term in the LME when using the GBLM as opposed to the linear regression approach. To the extent that GBLM analysis yielded a high level of rule-based knowledge, low residuals for the ecological model, but high residuals for the human judge, this second term would be low (due to lack of a high correlation between these residuals), and all signs would point to a person correctly following a set of rules, but doing so in a noisy fashion. Another interesting case would arise should GBLM analysis yield a high value for the second LME term: one interpretation would be that both the ecology and the human’s judgment strategy both incorporated either linear or continuous (rather than categorical) cue reliance, a signal that perhaps the standard Lens Model rather than the GBLM may provide a more plausible description of the judgment system.

Finally, the possibility may also exist that one type of model (compensatory or noncompensatory) provides the best fit to the environment, whereas the opposing model provides the best fit to the human’s judgment strategy. A situation such as this may raise even more interesting challenges and issues in achieving psychologically plausible interpretations for the parameters of the LME resulting from such a case. As we hope to have demonstrated in this article, we believe that a toolbox comprised of techniques for inferring *both* compensatory and noncompensatory judgment strategies and descriptions of environmental structure gives rise to an almost unlimited set of potentially interesting research questions and opportunities, both theoretical and empirical, in the analysis and modeling of human learning and performance.

REFERENCES

- Bisantz, A. M., Kirlik, A., Gay, P., Phipps, D., Walker, N., & Fisk, A. D. (2000). Modeling and analysis of a dynamic judgment task using a lens model approach. *IEEE Transaction on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30, 605-616.
- Bisantz, A. M., Llinas, J., & Drury, C. G. (2001). Issues in aided adversarial decision-making performance. In M. McNeese, E. Salas, & M. Endsley (Eds.), *New trends in cooperative activities: Understanding system dynamics in complex environments* (pp. 191-204). Santa Monica: Human Factors and Ergonomics Society.
- Brehmer, A. & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C. R. B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). Amsterdam: North Holland.

- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193-217.
- Cannon-Bowers, J. A. & Salas, E. (1998). *Making decisions under stress: Implications for individual and team training* (p. 447). Washington, DC: American Psychological Association.
- Chen, H., Shankaranarayanan, G., She, L., & Iyer, A. (1998). A machine learning approach to inductive query by examples: An experiment using relevant feedback, id3, genetic algorithms, and simulated annealing. *Journal of the American Society for Information Science*, 49, 693-705.
- Cohen, M. S., Freeman, J. T., & Thompson, B. B. (1997). Training the naturalistic decision maker. In C. Zsombok & G. Klein (Eds.), *Naturalistic decision making* (pp. 257-268). Mahwah, NJ: Lawrence Erlbaum Associates.
- Collyer, S. C. & Malecki, G. S. (1998). Tactical decision making under stress: history and overview. In J. Cannon-Bowers & E. Salas (Eds.), *Decision making under stress: Implications for individual and team training* (pp. 3-15). Washington, DC: American Psychological Association Press.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego: Academic Press, Inc.
- Dawes, R. M. (1964). Social selection based on multidimensional criteria. *Journal of Abnormal and Social Psychology*, 68, 104-109.
- Dawes, R. M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.
- DeJong, K. A. & Spears, W. M. (1990). An analysis of the interacting roles of population size and crossover in genetic algorithms. Presented at First International Conference on Parallel Problem Solving from Nature. Dortmund, Germany.
- DeJong, K. A., Spears, W. M., & Gordon, D. F. (1993). Using genetic algorithms for concept learning. *Machine Learning*, 13, 161-188.
- Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychological Bulletin*, 73, 221-230.
- Fogarty, W. M. (1988). *Formal investigation into the circumstances surrounding the downing of a commercial airliner by the U.S.S. Vincennes (CG 49) on 3 July 1988*. Unclassified Letter to Commander in Chief, U.S. Central Command Ser. 1320.
- Ganzach, Y. & Czaczkes, B. (1995). On detecting nonlinear noncompensatory judgment strategies: Comparison of alternative regression models. *Organizational Behavior and Human Decision Processes*, 61, 168-176.
- Gigerenzer, G. & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650-669.

- Gigerenzer, G. & Kurz, E. M. (in press). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. Stewart (Eds.), *The essential Brunswik*. Oxford: Cambridge University Press.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. New York: Addison-Wesley Publishing.
- Greene, D. P. & Smith, S. F. (1993). Competition-based induction of decision models from examples. *Machine Learning*, 13, 229-257.
- Hammond, K. R. (1955). Probabilistic functionalism and the clinical method. *Psychological Review*, 62, 255-262.
- Hammond, K. R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Its past, present, future. In G. A. Klein, J. Orasanu, R. Calderwood, & C. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 205-227). Norwood, NJ: Ablex Publishing Corp.
- Hammond, K. R., Hamm, R. M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems Man and Cybernetics*, 17, 753-770.
- Hammond, K. R. & Summers, D. (1972). Cognitive control. *Psychological Review*, 79, 58-67.
- Hodge, K. A. (1997). Training for decision making in a complex, dynamic task environment: An empirical investigation focusing on part-task training and feedback. In *School of Psychology*. Atlanta: Georgia Institute of Technology.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Ann Arbor, MI: University of Michigan Press.
- Holland, J. H., Holyoak, K. F., Nisbett, R. E., & Thagard, P. R. (1986). *Induction*. Cambridge, MA: MIT Press.
- Horiguchi, Y., Sawaragi, T., & Akashi, G. (2000). Naturalistic human-robot collaboration based upon mixed-initiative interactions in teleoperating environment. Presented at IEEE International Conference on Systems, Man and Cybernetics. Nashville.
- Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review*, 71, 42-60.
- Kaempf, G. L., Klein, G. A., Thordsen, M. L., & Wolf, S. (1996). Decision making in complex command-and-control environments. *Human Factors*, 38, 220-231.
- Kirlik, A. (1995). Requirements for psychological models to support design: Toward ecological task analysis. In J. M. Flach, P. Hancock, J. Caird, & K. J. Vicente (Eds.), *Global perspectives on the ecology of human-machine systems* (pp. 68-120). Hillsdale, NJ: Erlbaum.

- Kirlik, A. & Bisantz, A. (1999). Cognition in human-machine systems: Experiential and environmental aspects of adaptation. In P. Hancock Ed.), *Handbook of perception and cognition: Human performance and ergonomics* (pp. 47-68). San Diego, CA: Academic Press.
- Kirlik, A., Fisk, A. D., Walker, N., & Rothrock, L. (1998). Feedback augmentation and part-task practice in training dynamic decision-making skills. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: implications for individual and team training* (pp. 91-113). Washington, DC: American Psychological Association.
- Kirlik, A., Miller, R. A., & Jagacinski, R. J. (1993). Supervisory control in a dynamic uncertain environment: A process model of skilled human-environment interaction. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 929-952.
- Kirlik, A., Walker, N., Fisk, A. D., & Nagel, K. (1996). Supporting perception in the service of dynamic decision making. *Human Factors*, 38, 288-299.
- Klein, G. A. (1999). Applied decision making. In P. A. Hancock (Ed.), *Handbook of perception and cognition: Human performance and ergonomics* (pp. 87-108). New York: Academic Press.
- Klein, G. A., Orasanu, J., Calderwood, R., & Zsombok, C. E. (1993). *Decision making in action: Models and methods* (p. 480). Norwood, NJ: Ablex Corporation.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Presented at International Joint Conference on Artificial Intelligence. Montreal, Quebec.
- Kozlowski, S., Toney, R. J., Mullins, M. E., Weissbein, D. A., Brown, K. G., & Bell, B. S. (2001). Developing adaptability: A theory for the design of integrated-embedded training systems. In E. Salas (Ed.), *Advances in human performance and cognitive engineering research*, Vol. 1 (pp. 59-124). New York: Elsevier Science.
- Mendelson, E. (1997). *Introduction to mathematical logic* (4th ed.). London: Chapman & Hall.
- Michalski, R. S. (1983). A theory and methodology of inductive learning. In R. S. Michalski, J. G. Carbonell, & T. M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach*. Palo Alto, CA: Tioga Publishing Co.
- Orasanu, J. & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp. 3-20). Norwood, NJ: Ablex.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 16, 366-387.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *JEP: Learning, Memory, and Cognition*, 14, 534-552.

- Pearl, J. (2000). *Causality: Models, reasoning, and inference*. New York: Cambridge University Press.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1, 81-106.
- Rasmussen, J. R. (1983). Skills, rules and knowledge: signals, signs, symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 257-266.
- Reason, J. (1987). Cognitive aids in process environments: Prostheses or tools? *International Journal of Man-Machine Studies*, 27(5-6), 463-470.
- Rothrock, L. (1995). Performance measures and outcome analyses of dynamic decision making in real-time supervisory control. In *Industrial and Systems Engineering*. Atlanta: School of Industrial and Systems Engineering, Georgia Institute of Technology.
- Rothrock, L. & Repperger, D. W. (in review). The application of multi-objective genetic algorithms to design and validate human-machine interfaces in a manual control task. *IEEE Transactions on Systems, Man, and Cybernetics*.
- Rouse, W. B. (1983). Models of human problem solving: detection, diagnosis, and compensation for system failures. *Automatica*, 19, 613-625.
- Sawaragi, T., Horiguchi, Y., & Ishizuka, M. (2001). Comparative analysis of ecological expert and fuzzy controller using lens model. Presented at 10th IEEE International Conference on Fuzzy Systems. Melbourne, Australia.
- Sheridan, T. B., & Johanssen, G. (1976). Monitoring behavior and supervisory control. In *NATO Conference Series*. New York: Plenum Press.
- Smith, S. F. (1983). Flexible learning of problem solving heuristics through adaptive search. Presented at International Joint Conference on Artificial Intelligence. Karlsruhe, GE.
- Tucker, L. R. (1964). Suggested alternative formulation in the developments by Hursch, Hammond, and Hursch, and by Hammond, Hursch, and Todd. *Psychological Review*, 71, 528-530.
- Vafaie, H., & DeJong, K. (1994). Improving a rule induction system using genetic algorithms. In R. Michalski & G. Tecuci (Eds.), *Machine learning: A multistrategy approach*, Vol. IV. San Mateo, CA: Morgan Kaufman Publishers.
- Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of evidence. *Journal of Applied Psychology*, 59, 555-561.